Durham University

Durham Research Online

Deposited in DRO:

20 August 2021

Version of attached file:

Published Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Jiménez, Esteban and Padilla, Nelson and Contreras, Sergio and Zehavi, Idit and Baugh, Carlton M and Orsi, Álvaro (2021) 'The assembly bias of emission-line galaxies.', Monthly Notices of the Royal Astronomical Society, 506 (3).

Further information on publisher's website:

https://doi.org/10.1093/mnras/stab1819

Publisher's copyright statement:

This article has been accepted for publication in Monthly notices of the Royal Astronomical Society. ©: 2020 The Author(s). Published by Oxford University Press on behalf of the Royal Astronomical Society. All rights reserved.

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full DRO policy for further details.

Durham University Library, Stockton Road, Durham DH1 3LY, United Kingdom
Tel : +44 (0)191 334 3042 | Fax : +44 (0)191 334 2971
https://dro.dur.ac.uk

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **506,** 3155–3168 (2021)
Advance Access publication 2021 July 2

https://doi.org/10.1093/mnras/stab1819

The assembly bias of emission-line galaxies

Esteban Jiménez [1,2★] Nelson Padilla [1,3] Sergio Contreras,[4] Idit Zehavi,[5] Carlton M. Baugh [6] and Álvaro Orsi[7,8]

[1]*Instituto de Astrofísica, Pontificia Universidad Católica de Chile, 4860 Santiago, Chile*
[2]*International Centre for Radio Astronomy Research (ICRAR), University of Western Australia, Crawley, WA 6009, Australia*
[3]*Centro de Astro-Ingeniería, Pontificia Universidad Católica de Chile, 4860 Santiago, Chile*
[4]*Donostia International Physics Center (DIPC), Manuel Lardizabal Pasealekua 4, E-20018 Donostia, Basque Country, Spain*
[5]*Department of Physics, Case Western Reserve University, Cleveland, OH 44106, USA*
[6]*Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*
[7]*Centro de Estudios de Física del Cosmos de Aragón, Plaza San Juan 1, Planta 2, E-44001 Teruel, Spain*
[8]*PlantTech Research Institute Limited, South British House, 4th Floor, 35 Grey Street, Tauranga 3110, New Zealand*

Accepted 2021 June 16. Received 2021 June 16; in original form 2020 October 16

ABSTRACT

The next generation of spectroscopic surveys will target emission-line galaxies (ELGs) to produce constraints on cosmological parameters. We study the large-scale structure traced by ELGs using a combination of a semi-analytical model of galaxy formation, a code that computes the nebular emission from H II regions using the properties of the interstellar medium, and a large-volume, high-resolution *N*-body simulation. We consider fixed number density samples where galaxies are selected by their H α, [O III] λ5007, or [O II] $\lambda\lambda$3727–3729 emission-line luminosities. We investigate the assembly bias signatures of these samples, and compare them to those of stellar mass- and star formation rate-selected samples. Interestingly, we find that the [O III]- and [O II]-selected samples display scale-dependent bias on large scales and that their assembly bias signatures are also scale dependent. Both these effects are more pronounced for lower number density samples. The [O III] and [O II] emitters that contribute most to the scale dependence tend to have a low gas-phase metallicity and are preferentially found in low-density regions. We also measure the baryon acoustic oscillation (BAO) feature and the β parameter related to the growth rate of overdensities. We find that the scale of the BAO peak is roughly the same for all selections and that β is scale dependent at large scales. Our results suggest that ELG samples include environmental effects that should be modelled in order to remove potential systematic errors that could affect the estimation of cosmological parameters.

Key words: galaxies: evolution – galaxies: formation – galaxies: statistic – large-scale structure of Universe.

1 INTRODUCTION

Mapping the Universe using photometric and spectroscopic surveys allows us to measure the cosmic large-scale structure that encodes valuable cosmological information. However, since the galaxy distribution is not a direct tracer of the underlying density field it is essential to understand the connection between galaxies and dark matter haloes to obtain an accurate interpretation of the Universe (for a review, see Wechsler & Tinker 2018). A relevant statistical property of the galaxy distribution is the clustering signal that is often quantified using the two-point correlation function (2PCF). Measuring galaxy clustering allows us to extract two pieces of information that can be used to constrain the cosmological model (Weinberg et al. 2013): (i) the scale of standard ruler features, such as the baryonic acoustic oscillation (BAO) peak, from which we can obtain the cosmic expansion history, and (ii) the magnitude of redshift-space distortions (RSDs) in the clustering signal, which are driven by the rate at which structure grows. Both of these quantities depend on the amount of dark matter and dark energy.

Precise measurements of BAO and RSD are difficult to obtain because cosmic (sample) variance can be significant when the sampled volume is small (Kaiser 1986a) and shot noise can also reduce the signal-to-noise of a clustering estimate. The Sloan Digital Sky Survey (SDSS) and the 2dF Galaxy Redshift Survey were the first to observe hundreds of thousands of galaxies in large volumes, obtaining convincing detections of the BAO (Cole et al. 2005; Eisenstein et al. 2005). Observations have continued throughout the last 15 yr mostly using massive luminous red galaxies (LRGs) as the tracers of the large-scale structure (e.g. Eisenstein et al. 2011; Zehavi et al. 2011; Dawson et al. 2013; Bautista et al. 2018).

Advances in wide field spectroscopy have opened up the opportunity to trace the large-scale structure using emission-line galaxies (ELGs). The nebular emission of these galaxies is produced by gas in H II regions that is photoionized by radiation from young stars. Some emission-line luminosities have therefore been used to infer star formation rates (SFRs), although in general the relation between line emission and SFR can be complicated as the emission depends on the local properties of the ISM such as gas metallicity, temperature, ionization state, and density (e.g. Levesque, Kewley & Larson 2010; Gutkin, Charlot & Bruzual 2016; Byler et al. 2017) and on the

★ E-mail: esteban.jimenez@icrar.org

© 2021 The Author(s)
Published by Oxford University Press on behalf of Royal Astronomical Society

Downloaded from https://academic.oup.com/mnras/article/506/3/3155/6313299 by Durham University Library user on 19 August 2021

attenuation by dust of the line luminosity. Moreover, ELGs do not trace the field in the same way as LRGs; ELGs tend to reside in low-mass haloes (Favole et al. 2016; Gonzalez-Perez et al. 2018) and live in filaments and sheets rather than in the knots of the cosmic web occupied by LRGs (Gonzalez-Perez et al. 2020).

Typically, given the depth of upcoming surveys with spectrographs that operate in the optical, ELG catalogues have redshift distributions that peak around $z \sim 1$, which, in combination with Ly α emitters at $z \gtrsim 2$, enables us to investigate the history of cosmic expansion at previously unexplored epochs. The SDSS-IV/eBOSS survey (Alam et al. 2021) provides one of the largest ELG catalogues to date, and the next generation of surveys like DESI (DESI Collaboration 2016) and *Euclid* (Laureijs et al. 2011) will detect millions of ELGs. This will potentially enable us to measure cosmological parameters with exceptional precision.

To fully exploit the future ELG data, it is essential to understand any systematic effects that may influence the inferred cosmological constraints. Galaxy formation models can be used for this purpose, to provide insights into the galaxy–halo connection with the by-product of testing different prescriptions for the physical processes that regulate galaxy evolution. A useful approach to explore galaxy formation is to use semi-analytical models (SAMs; Cole et al. 2000; Baugh 2006; Somerville et al. 2008). These models use simplified descriptions of physical processes that shape the fate of baryons within the dark matter halo merger trees extracted from *N*-body dark matter-only simulations, expressed in a set of coupled differential equations with parameters to encapsulate 'subgrid' physics. SAMs can successfully reproduce, among other things, the observed luminosity and stellar mass functions (e.g. Croton et al. 2006; Henriques et al. 2015; Croton et al. 2016; Stevens, Croton & Mutch 2016; Lagos et al. 2018; Baugh et al. 2019). Alternatively, hydrodynamical simulations offer a complementary approach to follow baryonic physics, which in general requires fewer assumptions and approximations than are needed in SAMs, but which nevertheless still appeal to subgrid recipes for unresolved processes (e.g. Vogelsberger et al. 2014; Schaye et al. 2015; Nelson et al. 2018). Due to the higher computational overhead of hydrodynamical simulations compared with SAMs, the largest volumes probed by state-of-the-art hydrodynamical simulations are still 10–100 times smaller than the typical SAM volume.

Another approach used to connect galaxies with their host haloes is to employ an empirical model such as the halo occupation distribution (HOD) framework (e.g. Benson et al. 2000; Scoccimarro et al. 2001; Berlind & Weinberg 2002; Kravtsov et al. 2004; Zheng et al. 2005). The HOD provides an empirical relation between the average number of galaxies N hosted by haloes with mass M. This relation is characterized by a probability distribution $P(N|M)$ that depends on the redshift, number density, and selection criteria of a galaxy sample. Here, the standard assumption is that the galaxy content depends only on halo mass, but this may not be true if the galaxy distribution correlates with the assembly history of haloes. *N*-body simulations have shown that the clustering of dark matter haloes does depend on secondary halo properties like formation time, concentration, and spin (e.g. Gao, Springel & White 2005; Wechsler et al. 2006; Gao & White 2007), an effect called halo assembly bias. Likewise, the manifestation of assembly bias in galaxy clustering, commonly referred to as galaxy assembly bias, has been found both in SAMs (e.g. Croton, Gao & White 2007; Zehavi et al. 2018, 2019; Contreras et al. 2019; Jiménez et al. 2019; Xu, Zehavi & Contreras 2021) and hydrodynamical simulations (e.g. Artale et al. 2018; Montero-Dorta et al. 2020; Xu & Zheng 2020). Observational evidence for assembly bias remains inconclusive; some studies have suggested detections (e.g. Lacerna, Padilla & Stasyszyn 2014; Hearin, Watson & van den Bosch 2015; Watson et al. 2015; Montero-Dorta et al. 2017), while others indicate that the imprint on the galaxy distribution is small (e.g. Lin et al. 2016; Zu & Mandelbaum 2016; Walsh & Tinker 2019). Assessing the existence of assembly bias in the real Universe is an important task; cosmological constraints from future surveys will most likely be limited by how well we can model the observations rather than the precision of the measurements.

In general, assembly bias enhances the clustering amplitude on large scales for stellar mass-selected samples and suppresses it for SFR selections (Contreras et al. 2019; Contreras, Angulo & Zennaro 2021). However, as we found in these studies, these trends can change depending on the number density and redshift of the sample. So far, there are no direct measurements of the assembly bias signature in ELG catalogues, but, in principle, the effect should be similar to that reported for SFR selections as star-forming ELGs are a subset of SFR selections.

Here, we aim to study the large-scale structure of ELGs by measuring the clustering and galaxy assembly bias signature of these samples. Note that ELGs can be both star-forming galaxies and active galactic nuclei (AGNs). We focus on star-forming ELGs here and do not model AGN emission lines. We employ galaxies from the SAG SAM (Cora et al. 2018) run on the MultiDark Planck cosmological simulation (Klypin et al. 2016). The total simulated volume is $(1\ h^{-1}\,\mathrm{Gpc})^3$, so the effect of the sample variance is greatly reduced. Thus, we can accurately sample the 2PCF up to scales of the BAO feature and determine whether or not the impact of assembly bias from ELG selections is significant. We calculate the nebular emission in SAG galaxies using the GET_EMLINES code from Orsi et al. (2014), and then store the H α, [O III], and [O II] line emission luminosities. These emission lines at $z \sim 1$ correspond to the near-infrared and optical range sampled by DESI and *Euclid*, respectively.

The layout of the paper is as follows: Section 2 describes the SAG galaxy formation model and the *N*-body simulation in which it is implemented, along with how we define our galaxy samples, while in Section 3 we compare and characterize these samples. In Section 4, we show the assembly bias signatures, and in Section 5 we study a possible origin for features in the assembly bias of ELGs. The BAO and the β parameter, which quantifies the strength of anisotropies produced by the RSD, are shown in Section 6 for each sample. We conclude in Section 7. Brief discussions about results from other SAMs, and ELG sample completeness, are presented in Appendixes A and B, respectively.

Throughout the paper, masses are measured in $h^{-1}\,\mathrm{M}_\odot$, the SFR is measured in $h^{-1}\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$, and distances are measured in h^{-1} Mpc and are in comoving units.

2 SIMULATION DATA

2.1 The galaxy formation model: SAG

Here, we use the Semi-Analytical Galaxies (SAG) model of galaxy formation (Cora 2006). SAMs use the merger trees extracted from *N*-body simulations to model the main physical processes involved in the evolution of a galaxy, such as the treatment of radiative cooling of hot gas, star formation, feedback effects from supernovae and AGN, chemical enrichment of the gas, the growth of supermassive black holes, and the impact from galaxy mergers, among others.

The version of SAG used here is the one presented in Cora et al. (2018), which uses the outputs of the MultiDark2 cosmological simulation (MDPL2; Klypin et al. 2016; see Section 2.2 for more

Downloaded from https://academic.oup.com/mnras/article/506/3/3155/6313299 by Durham University Library user on 19 August 2021

details). The main output of the simulation and the SAM are publicly available[1] as a part of the MultiDark comparison project (Knebe et al. 2018). The SAG SAM was originally presented in Cora (2006) and is based on the model by Springel et al. (2001). Since then, the code has been through several updates (Lagos, Cora & Padilla 2008; Tecce et al. 2010; Orsi et al. 2014; Padilla et al. 2014; Gargiulo et al. 2015; Muñoz Arancibia et al. 2015) and is capable of reproducing observations at both low and high redshifts. One of the key features of this model is the use of the particle swarm optimization technique to automatically set the model parameters by requiring the output to fit several observables (Ruiz et al. 2015).

The galaxy properties produced by this model include stellar mass, cold gas, black hole and bulge masses, the average and instantaneous SFRs, where the latter corresponds to the SFR in the most recent time substep, which is a subdivision of the time-step between the simulation snapshots. Substep sizes range from 5 and 15 Myr, whereas the time between snapshots is of the order of 100 Myr. These quantities are computed separately for discs and bulges, where the former are the result of quiescent star formation in cooled gas discs and the latter form via starburst episodes. The gas-phase metallicity for both components is calculated by modelling the chemical enrichment of the ISM, which takes into account mass-loss from massive stars and supernovae. These two ways to compute the SFR are important when computing the emission-line fluxes of the galaxies (see Section 2.3 for more details).

2.2 The MultiDark Planck 2 simulation

As mentioned in the previous section, SAG was run on the halo merger trees from the MULTIDARK simulation MDPL2 (Klypin et al. 2016). The MDPL2 adopts a ΛCDM Universe, characterized by Planck cosmological parameters (Planck Collaboration XVI 2014): $\Omega_{\rm m} = 0.307$, $\Omega_{\rm b} = 0.048$, $\Omega_\Lambda = 0.693$, $h = 0.678$, and $n_{\rm s} = 0.96$. The simulation follows 3840^3 particles within a cubic box of comoving side length 1 h^{-1} Gpc, with a mass resolution of $m_{\rm p} = 1.51 \times 10^9\, h^{-1}\, {\rm M}_\odot$. The particles are followed from $z = 120$ until $z = 0$ and their positions and velocities are output at 126 snapshots. The dark matter haloes are identified with the ROCKSTAR halo finder (Behroozi, Wechsler & Wu 2013a), and the CONSISTENT TREES code (Behroozi et al. 2013b) is used to construct the merger trees. These halo finder and halo merger tree algorithms identify objects in phase space, keeping a better track of the substructures after their infall.

The large cosmological volume of the MDPL2 allows us to make accurate clustering measurements up to scale separations that encapsulate useful cosmological information.

2.3 The calculation of nebular emission

To model the nebular emission of the star-forming galaxies, we use the GET_EMLINES model[2] introduced by Orsi et al. (2014) (hereafter O14) to post-process the output from the SAG model. The GET_EMLINES code uses the output of the photoionization code MAPPINGS-III (Dopita & Sutherland 1995; Groves, Dopita & Sutherland 2004), as tabulated by Levesque et al. (2010). MAPPINGS-III predicts the nebular emission from H II regions. The grid calculated by Levesque et al. (2010) tabulates the emission-line fluxes as a function of the gas-phase metallicity and the ionization parameter of the H II region, q. O14 use the gas metallicities of the bulges and discs from the SAG galaxies, but it does not predict the ionization parameter within individual H II regions due to a lack of resolution to resolve the internal structure of galaxies. Instead, O14 advocated a model in which q could be inferred from the gas-phase metallicity, an assumption that is inspired by observational results that suggest that q is anticorrelated with the metallicity of the cold star-forming gas (e.g. Nagao, Maiolino & Marconi 2006; Groves & Allen 2010; Shim & Chary 2013). O14 show that parameters selected in their model to calculate q from the gas-phase metallicity allowed the SAG model to reproduce the locus of star-forming galaxies in the so-called BPT diagram relating the line ratios [O III λ5007]/H β and [N II λ6854]/H α (Baldwin, Phillips & Terlevich 1981). Note that O14 showed that the predictions were robust to substantial perturbations to the parameter values in the model for q. Ideally, the GET_EMLINES code uses the instantaneous SFR rather than a time-averaged SFR, as the instantaneous SFR is a better indicator of the number of Lyman continuum photons that make up the ionizing radiation field and, as a consequence, of the H α luminosity. Nevertheless, Favole et al. (2020) used the averaged SFR predicted by the SAGE (Croton et al. 2016) and GALACTICUS (Benson 2012) SAMs to infer the [O II] line emission, and found reasonable agreement with observational data at $z \sim 1$.

We use GET_EMLINES and instantaneous SFRs to compute the luminosities for the H α, [O III] λ5007, [O II] λλ3727–3729, and [N II] λλ6548–6584 emission lines (hereafter H α, [O III], [O II], and [N II], respectively). As these values are calculated separately for the bulge and disc components of a galaxy, we sum these contributions to obtain the total nebular emission for each galaxy. Note that we do not apply any attenuation to the line luminosity.

2.4 The galaxy samples

We use galaxy samples characterized by fixed number densities of objects. We achieve this by ranking the galaxies from the highest to lowest values of a given property (e.g. their emission-line luminosity, averaged SFR, or stellar mass), and then retaining only those galaxies above a threshold value that correspond to the desired number density. The number densities used in this work are $n = 0.001\, h^3\, {\rm Mpc}^{-3}$, $n = 0.003\,16\, h^3\, {\rm Mpc}^{-3}$, and $n = 0.01\, h^3\, {\rm Mpc}^{-3}$.

As the samples selected by line emission are star-forming galaxies, it is expected that they will have some overlap with galaxies selected by their SFR. Hence, we also include SFR-selected samples and perform the same analysis on these as carried out for the ELG samples. We also consider stellar mass-selected samples to allow further comparisons with assembly bias signatures that have been studied in several galaxy formation models (e.g. Artale et al. 2018; Zehavi et al. 2018; Contreras et al. 2019). The cumulative distribution functions for the different selections are shown in Fig. 1. The number densities adopted to define our samples are shown by the horizontal thin dashed lines. Note that the galaxies included in each sample are those to the right of the intersection between the cumulative functions with the horizontal lines. We also include a selection based on the combined luminosity of the H α+doublet [N II] λλ6548–6584 lines; this combination mimics the flux that the *Euclid* mission will capture for its sources due to its limited spectral resolution. This allows us to assess the contribution of the [N II] flux to the predictions for H α (see also Merson et al. 2018).

The data used here correspond to a subsample of SAG[3] where galaxies were selected with a stellar mass cut of $5 \times 10^8\ h^{-1}\, {\rm M}_\odot$.

[1] http://www.cosmosim.org
[2] https://github.com/aaorsi/get_emlines
[3] http://skiesanduniverses.org/

Downloaded from https://academic.oup.com/mnras/article/506/3/3155/6313299 by Durham University Library user on 19 August 2021

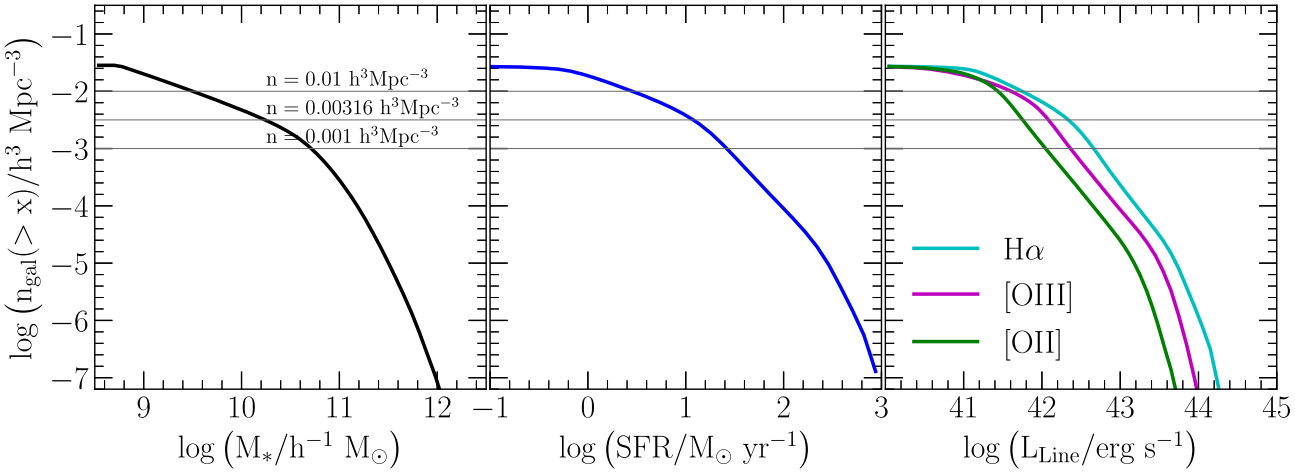

Figure 1. The cumulative distribution functions of SAG galaxies selected by stellar mass (*left*), SFR (*middle*), and Hα, [O III], and [O II] line luminosities (*right*). The horizontal dashed lines indicate the number densities used to define our galaxy samples. Note the plateau at low SFR and $L_{\rm Line}$, which is due to a stellar mass cut in the parent catalogue, which is discussed further in Appendix B.

This cut affects the completeness of SFR and ELG selections. In Appendix B, we explain that this has a negligible impact on the trends and results we obtain.

3 PROPERTIES OF ELG-SELECTED SAMPLES

To assess the level of similarity between the ELG and star-forming galaxy samples, we compare their 2PCFs. The 2PCF measures the excess probability of finding a pair of objects at a separation r compared to a homogeneously distributed sample. We measure the 2PCF using the CORRFUNC public code presented in Sinha & Garrison (2020).[4] The resulting 2PCFs are shown in the main panel of Fig. 2 for the SFR and ELG samples with a number density $n = 0.003\,16\ h^3\ {\rm Mpc}^{-3}$; the subpanel shows the ratios between these measurements and the 2PCF of the SFR-selected sample. The first impression is that the shapes of the 2PCFs are largely similar, irrespective of separation, with variations within 10 per cent. On large scales, the differences are mostly due to the different bias parameters of the samples, with the [O III]- and [O II]-selected samples showing weaker clustering than the Hα and SFR samples. On small scales, the differences may be due to the additional dependence on the physical conditions in the ISM; the Hα emission mostly traces the SFR, but the [O III] and [O II] emission also depends on the cold gas metallicity. Hence, differences in the one-halo terms suggest a possible connection between the spatial distribution of ELGs and the properties of their ISM. As can be seen from Fig. 2, the difference in the 2PCF compared with that measured for the SFR-selected sample depends on which line is chosen to construct the sample. Both the Hα and Hα + [N II] selections result in an amplitude for the 2PCF that is similar to that found for the SFR selection. This is expected as the strength of Hα emission is almost an instantaneous measure of the SFR, with little dependence on the metallicity of the star-forming gas. The 2PCFs measured for the [O III] and [O II] selections show a bigger difference from that found for the star formation sample, with an amplitude reduction of ∼20 and 30 per cent, respectively. This change in the effective bias parameter of these samples is related to the selection

[4]https://corrfunc.readthedocs.io/en/master/

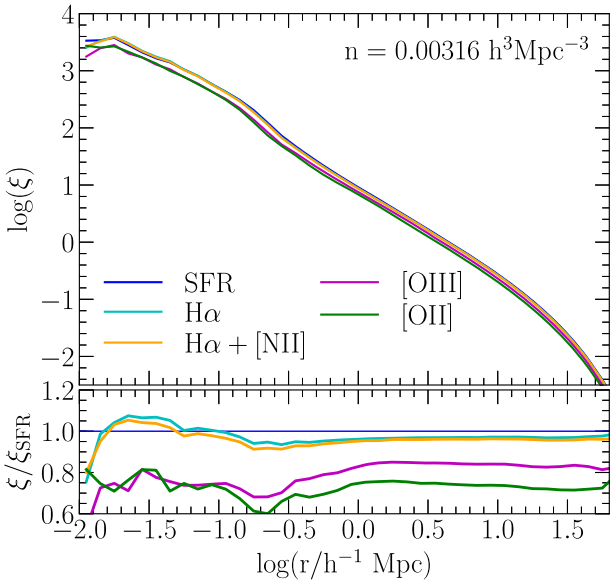

Figure 2. The 2PCFs for SAG samples selected according to different properties, as indicated by the key; in each case, the number density is $n = 0.003\,16\ h^3\ {\rm Mpc}^{-3}$. The bottom panel shows the ratio of the correlation functions measured for each sample with respect to the SFR-selected sample.

of galaxies with specific combinations of SFR and gas metallicity, as we demonstrate below.

One way to interpret the 2PCF is by using the HOD framework. The HOD characterizes a galaxy population via the halo occupation function, the average number of galaxies as a function of the host halo mass. Whereas HODs are typically used as an empirical model with parameters that are set to reproduce the measured abundance and 2PCF of a galaxy samples, SAMs predict the form of the HOD. So, in the case of SAMs, the HOD formalism produces a concise description of the model output that can be readily interpreted in relation to the 2PCF. In general, this function is separated into the contribution from central galaxies and satellites with specific forms that depend on the selection criteria (Zheng et al. 2005). For example, when samples are defined by luminosity or stellar mass cuts, the

Downloaded from https://academic.oup.com/mnras/article/506/3/3155/6313299 by Durham University Library user on 19 August 2021

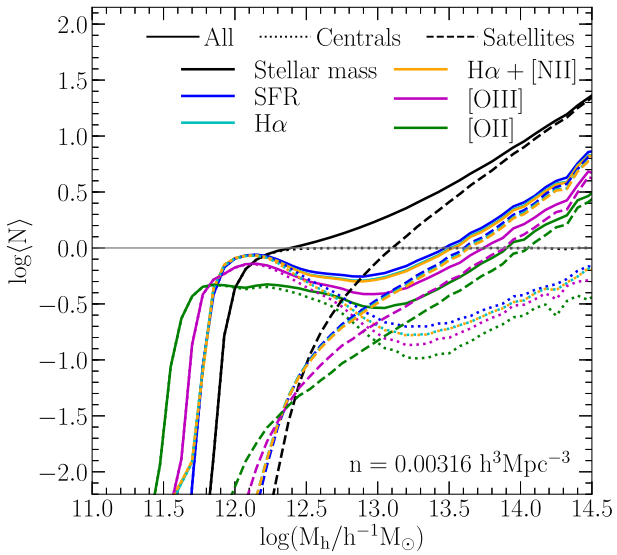

Figure 3. The HODs predicted by the SAG model for all galaxies (solid), centrals (dotted), and satellites (dashed), with different colours indicating different galaxy selections, as shown by the figure key. All samples have a number density of $n = 0.003\,16\ h^3\,\mathrm{Mpc}^{-3}$.

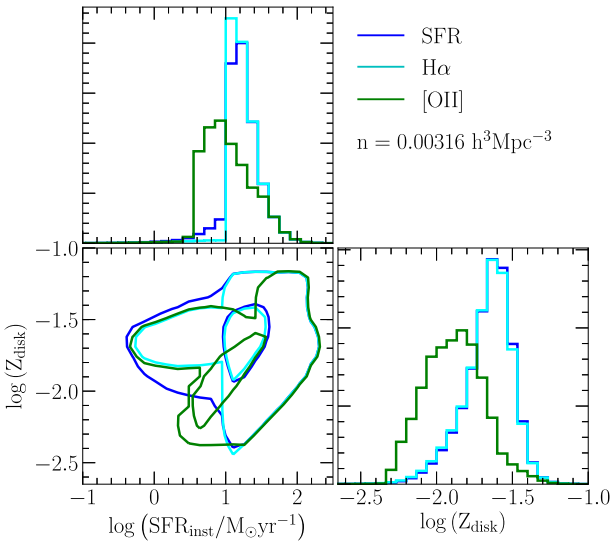

Figure 4. Distributions of instantaneous SFR ($\mathrm{SFR_{inst}}$) and cold gas metallicity of the disc ($\mathrm{Z_{disc}}$) for SAG samples with a fixed number density $n = 0.003\,16\ h^3\,\mathrm{Mpc}^{-3}$. Different colours correspond to different selections as shown by the key; the density contours enclose 68 per cent and 99 per cent of the distribution of galaxies. The corner panel shows the bivariate distributions, whereas the top and right-hand panels show the marginalized distributions of instantaneous SFR and cold gas metallicity, respectively.

HOD for central galaxies follows a step-like form. When samples are defined by SFR or colour cuts, on the other hand, the HOD of centrals reaches a peak followed by a dip to values below unity as the halo mass increases (e.g. Zehavi et al. 2011; Contreras et al. 2013; Gonzalez-Perez et al. 2018; Jiménez et al. 2019). The output of a SAM for these selections can be tabulated as an HOD, without having to adopt a particular parametric form, which is very powerful when considering selections for which there is little available data, such as ELGs.

Fig. 3 shows the HODs predicted by the SAG model for SFR and ELG selections with a number density $n = 0.003\,16\ h^3\,\mathrm{Mpc}^{-3}$. We also show results for a stellar mass-selected sample to illustrate the differences compared to the star-forming and ELG samples. The stellar mass-selected sample in the figure shows the canonical step-like form for the HODs of centrals, with the HOD for satellites exhibiting a power-law behaviour. In contrast, the ELG selections show a peak in the HOD of central galaxies, which shifts to lower masses for the [O III] and [O II] selections. Similar results have been recently found for eBOSS ELG samples (Guo et al. 2019; Avila et al. 2020). This indicates that model ELGs are mostly hosted by low halo masses, consistent with previous results from simulations (e.g. Gonzalez-Perez et al. 2018). For large halo masses, the HODs of both centrals and satellites increase with halo mass.

The overlap between the SFR- and ELG-selected samples can be quantified by analysing the similarities in their galaxy properties. Fig. 4 shows the distribution of galaxies in the instantaneous SFR versus gas metallicity plane for the SFR, Hα, and [O II] selections, in all cases with a number density $n = 0.003\,16\ h^3\,\mathrm{Mpc}^{-3}$. As can be seen, the distribution of the Hα-selected sample is in very good agreement with that of the SFR selection, as expected. In contrast, the distribution for the [O II]-selected sample is shifted to lower instantaneous SFR and cold gas metallicity. Still, an important fraction of [O II] emitters overlap with the SFR and Hα selections. Table 1 shows the fraction of overlap between the ELG and SFR selections. Note that ELG- and SFR-selected samples have less overlap at lower number densities.

Table 1. Fraction of SFR-selected galaxies that also satisfy the ELG selection criteria. Different rows indicate the number density of the samples as shown in the first column. Columns 2, 3, and 4 give the fraction of objects that also meet the Hα, [O III], and [O II] selections, respectively.

$n/h^3\,\mathrm{Mpc}^{-3}$	Hα	[O III]	[O II]
0.001	0.81	0.57	0.39
0.003 16	0.91	0.71	0.47
0.01	0.96	0.91	0.79

4 THE GALAXY ASSEMBLY BIAS OF ELG SAMPLES

Measurements from *N*-body simulations have shown that in order to fully determine the clustering of dark matter haloes one needs, in addition to their masses, knowledge of secondary halo properties such as formation time, concentration, subhalo occupation, and spin (e.g. Gao et al. 2005; Wechsler et al. 2006; Gao & White 2007). This effect, termed halo assembly bias, may potentially have an impact on the galaxy content of haloes, producing variations in the halo occupation functions and therefore affecting the large-scale galaxy clustering amplitude (e.g. Artale et al. 2018; Zehavi et al. 2018; Contreras et al. 2019). Hence, it is important to model this effect when interpreting the correlation function using the standard HOD framework.

SAM samples that are obtained using halo merger histories extracted from *N*-body simulations are affected by assembly bias because the growth histories of dark matter haloes, and therefore the level of assembly bias that haloes are subject to, also affect the galaxies that evolve within them. To measure the impact of assembly bias on the clustering of our galaxy samples, we compare their 2PCFs with that of 'shuffled' galaxy samples. The shuffling removes information about the assembly history of haloes by randomly

Downloaded from https://academic.oup.com/mnras/article/506/3/3155/6313299 by Durham University Library user on 19 August 2021

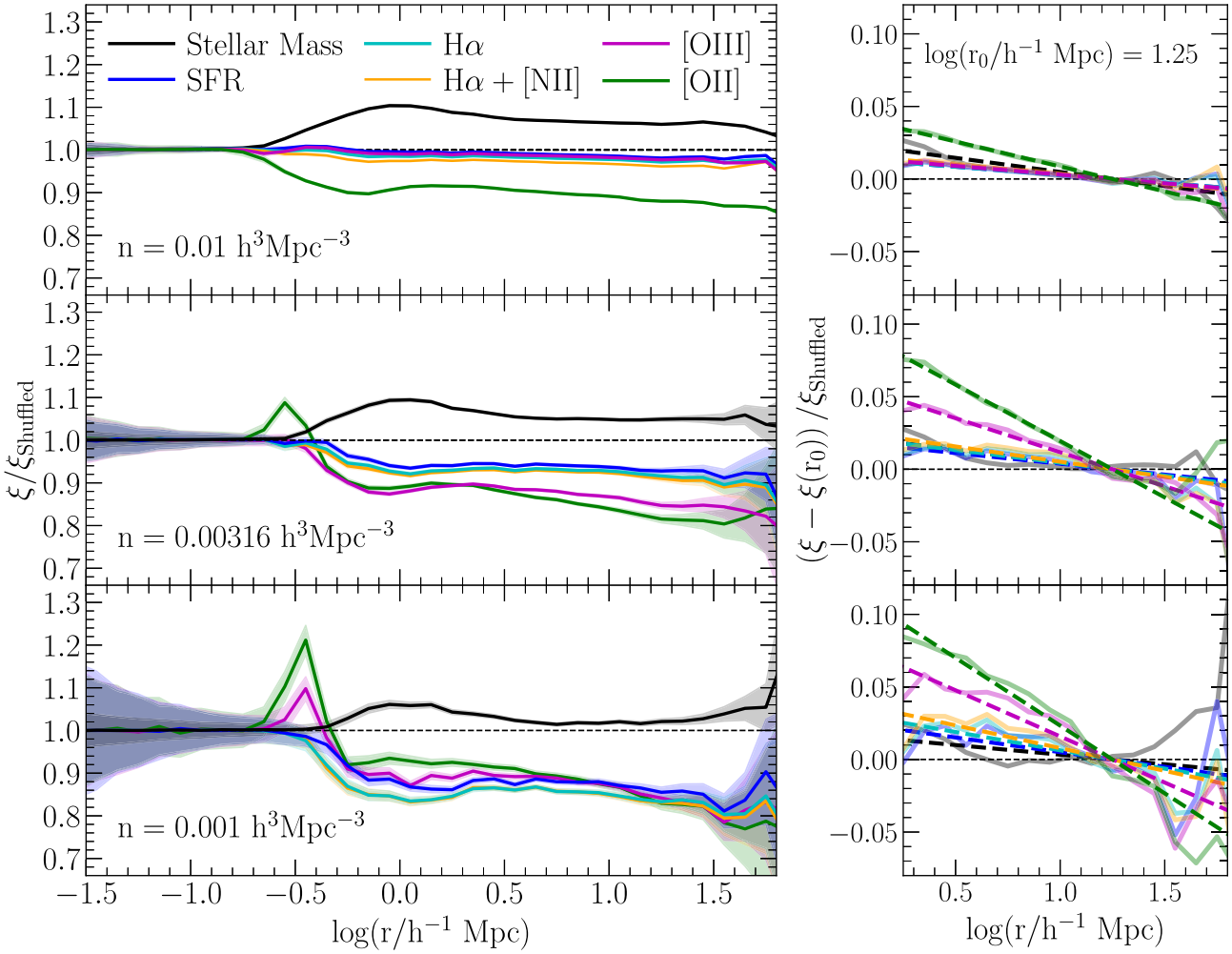

Figure 5. (*Left*) The assembly bias signature in the SAG samples. Each panel shows a different number density as labelled. Note that the assembly bias for the [O III] and [O II] selections exhibits a clear scale-dependent signature for the two lowest number densities. Shaded regions represent the errors estimated from 1000 jackknife realizations. (*Right*) Zoom-in views of the assembly bias signatures shown in the left-hand panels. Note that we have re-normalized the *Y*-axes for a clearer visualization of the slope differences. Dashed lines indicate the best-fitting linear functions for the assembly bias signatures in the $0.25 < \log(r/h^{-1}\,\mathrm{Mpc}) < 1.5$ range.

exchanging the galaxy populations between haloes of the same mass (Croton et al. 2007; Xu et al. 2021). The standard approach preserves the central–satellites distances; therefore, the one-halo terms of the shuffled catalogues are the same as those of the original SAM samples. The assembly bias signature can be obtained by comparing the 2PCF of the SAM samples to that of the shuffled samples.

The impact of assembly bias on galaxy clustering depends on the selection criteria, number density, and redshift of the sample (e.g. Contreras et al. 2019). As the ELG selection shows substantial overlap with selection by SFR (see Fig. 4), we can estimate how much of the effect of assembly bias on the clustering of ELGs comes from the SFR selection. We do this by looking at the assembly bias effect present in a purely SFR-selected sample. Even though nebular emission traces SFR, some properties of the gas in the ISM, such as metallicity, can introduce additional effects that are not included when selecting by SFR alone. The assembly bias signatures for SFR-, stellar mass-, and ELG-selected samples are shown in Fig. 5 for the three number densities. It can be seen from Fig. 5 that the assembly bias in the H α and H α + [N II] selections is similar to that seen in the SFR selection. In contrast, assembly bias suppresses the galaxy clustering of [O III] and [O II] selections by up to 30 per cent. Table 1 shows that the H α sample has a high overlap with the SFR-selected sample for all number densities considered. For [O III] emitters, the overlap with the SFR selection is high for the highest number density sample, explaining their similar clustering in the top panel of Fig. 5. For the other number densities considered, the overlap between the [O III]- and SFR-selected samples is much smaller and their clustering is different.

We estimate the assembly bias uncertainties using jackknife resampling (Norberg et al. 2009). The simulation volume is divided into 1000 equal cubes of side 100 h^{-1} Mpc. The resulting errors are shown as shaded regions in Fig. 5. The uncertainties are negligible for a wide range of large-scale separations, suggesting that the assembly bias signatures, including the scale-dependent feature for the [O III] and [O II] selections, are robust.

Furthermore, the assembly bias is scale dependent, particularly in the case of [O II]. This steepness in the ratio of the 2PCF to the shuffled samples is also present in the SFR, H α, and H α+[N II] selections but only for the lowest number densities and, in any case, it is not as scale dependent as in the [O III] and [O II] cases. Moreover, there is a 'bump' feature in the ratio that is present only for the latter two selections around the transition from the one-halo term

Downloaded from https://academic.oup.com/mnras/article/506/3/3155/6313299 by Durham University Library user on 19 August 2021

Table 2. The slopes of the assembly bias signatures, m, obtained from simple linear fits. Different rows and columns specify the selection and number density of each sample, respectively.

Abundance	$0.001\,h^3\,\mathrm{Mpc}^{-3}$	$0.003\,16\,h^3\,\mathrm{Mpc}^{-3}$	$0.01\,h^3\,\mathrm{Mpc}^{-3}$
Selection		m	
M_*	-0.013 ± 0.006	-0.018 ± 0.004	-0.019 ± 0.003
SFR	-0.020 ± 0.007	-0.015 ± 0.002	-0.011 ± 0.001
Hα	-0.025 ± 0.006	-0.018 ± 0.003	-0.011 ± 0.001
Hα + [N II]	-0.031 ± 0.007	-0.021 ± 0.003	-0.013 ± 0.001
[O III]	-0.064 ± 0.007	-0.047 ± 0.003	-0.012 ± 0.001
[O II]	-0.093 ± 0.005	-0.078 ± 0.002	-0.034 ± 0.001

to the two-halo term [$\log(r/h^{-1}\,\mathrm{Mpc}) \sim -0.4$]. Overall, there is a clear trend between number density and the impact of assembly bias on galaxy clustering; for all ELG selections, the suppression of the clustering amplitude is larger for samples with lower number densities (i.e. for galaxies with higher emission-line luminosities).

To quantify the steepness of the assembly bias signatures, we fit simple linear functions between $0.25 < \log(r/h^{-1}\,\mathrm{Mpc}) < 1.5$ scale separations. Note that we do not use the jackknife errors in the linear fitting as they are small and do not vary much with scale over this range of separations. Table 2 shows the fitted slopes for all selections and number densities. The right-hand panels of Fig. 5 show these fitting functions for each selection along with its assembly bias signature in the corresponding left-hand panel. For clarity, we have re-normalized the amplitude so that all cases match at $\log(r/h^{-1}\,\mathrm{Mpc}) = 1.25$.

The slopes fitted for the [O II] selections are ∼3.7, ∼4.3, and ∼3 times larger than those found for the Hα selections for the low, medium, and high number densities, respectively. Likewise, the [O III] selections exhibit slopes that are ∼2 times larger than the Hα selections. Note that the slopes for the stellar mass and SFR selections are lower than those from the [O III] or [O II] cases for the low and medium density samples. Furthermore, the steepness of the assembly bias signature in the [O III] and [O II] selections is more pronounced for lower number densities. We also show the impact on the stellar mass-selected samples; it can be seen that assembly bias enhances the galaxy clustering and shows no strong scale dependence for any of the number densities explored here.

To investigate the origin of the steepness in the assembly bias signature, we analyse the galaxy clustering of the SAG and shuffled samples separately, by comparing their 2PCFs to that of the dark matter. The dark matter 2PCF is obtained from the linear power spectra used in the MDPL2 simulation, which is Fourier transformed to obtain the linear theory matter correlation function. Using the 2PCF of the dark matter, $\xi_{\rm mm}$, and the 2PCF of the galaxy sample, $\xi_{\rm gal}$, we compute the large-scale bias of each sample using

$$b(r) = \sqrt{\frac{\xi_{\rm gal}}{\xi_{\rm mm}}}. \quad (1)$$

The value of the bias parameter is expected to be constant to first order on linear scales and to vary with the galaxy selection.

The main panels of Fig. 6 show the 2PCFs of the dark matter and the SAG and shuffled samples for the Hα, [O III], and [O II] selections, for the number density of $n = 0.003\,16\,h^3\,\mathrm{Mpc}^{-3}$. For clarity, we show results for the two-halo term only [$\log(r/h^{-1}\,\mathrm{Mpc}) \gtrsim 0.5$]. The estimate of the bias parameter, $b(r)$, for each sample is shown as a coloured line in the bottom panels. We average these values between $1.4 < \log(r/h^{-1}\,\mathrm{Mpc}) < 1.8$ (where the jackknife errors are small and do not vary strongly with scale) to obtain a constant large-scale bias, which is shown by the grey horizontal lines for comparison with $b(r)$. For the Hα selection, the bias parameter is roughly constant over the range of interest, for both the SAG and shuffled samples. In contrast, the bias parameters for the SAG [O III] and [O II] selections show a scale dependence. The bias for the shuffled samples is seen to be roughly constant. For the lowest number density sample (not shown here), we find that the bias parameter has an even steeper scale dependence for [O II] and [O III] selections. The larger value found for the bias of the Hα-selected sample indicates that, in this case, galaxies trace higher peaks in the density field than galaxies in the other ELG-selected samples.

The prediction of a scale-dependent bias parameter for the SAG [O III]- and [O II]-selected samples indicates that there are additional features that shape the large-scale clustering of these tracers. This suggests that the gas metallicity, which has an impact on the [O II] and [O III] emission for a given amount of star formation, has some dependence on environment. This is confirmed by the much weaker scale dependence found for the bias on large scales in the shuffled samples.

5 ORIGIN OF SCALE-DEPENDENT ASSEMBLY BIAS

In this section, we investigate the origin of the scale dependence of the galaxy assembly bias signature in galaxies selected by their [O III] and [O II] line emission. As just noted, the scale-dependent bias is only present in the original SAG samples and not in their shuffled counterparts. This suggests that the preference for the environment that characterizes galaxies with strong line emission, which could cause this scale dependence, is removed when shuffling these samples. Gonzalez-Perez et al. (2020) analysed how model ELGs trace the large-scale structure in an N-body simulation. They found that about half of [O II] emitters live in filaments while one-third live in sheets. This indicates that the galaxies selected using [O II] line luminosity will be preferentially located in low-density regions. Hence, quantifying the effect that the shuffling procedure has of moving ELGs to random locations in the cosmic web can provide an insight into the relation between the environment of [O II] emitters and the scale-dependent assembly bias. While there is some overlap between the SFR- and [O II]-selected samples, the SFR-selected sample shows little sign of scale-dependent bias. Hence, the source of the scale dependence is likely to be found in the galaxies that are not in common between the two samples. Indeed, Table 1 shows that the overlap between the SFR and [O II] emitters is less than 50 per cent for the two lowest density samples. This indicates that the origin of the scale dependence is encoded in the selection by [O II] luminosity.

One approach to quantifying the environment of a galaxy sample is to compute the local number density of main haloes around each galaxy. We use the main haloes as we associate galaxies with their host dark matter haloes rather than with subhaloes; not all of the satellite galaxies may be associated with a resolved subhalo. We define the number density, $n_{\rm local}$, using the distance to the fifth nearest main halo in the MDPL2 simulation, r_5, as $n_{\rm local} = 5/V(r_5)$, where $V(r_5)$ is the volume of a sphere of radius $r = r_5$. To count neighbouring haloes, we use those with masses $M_{\rm h} > 10^{10.8}\,h^{-1}\,\mathrm{M}_\odot$ for all galaxy selections. Note that we aim to quantify the change of galaxy environments (based on density) rather than determine a precise variation of the cosmic web environment after applying the shuffling

Downloaded from https://academic.oup.com/mnras/article/506/3/3155/6313299 by Durham University Library user on 19 August 2021

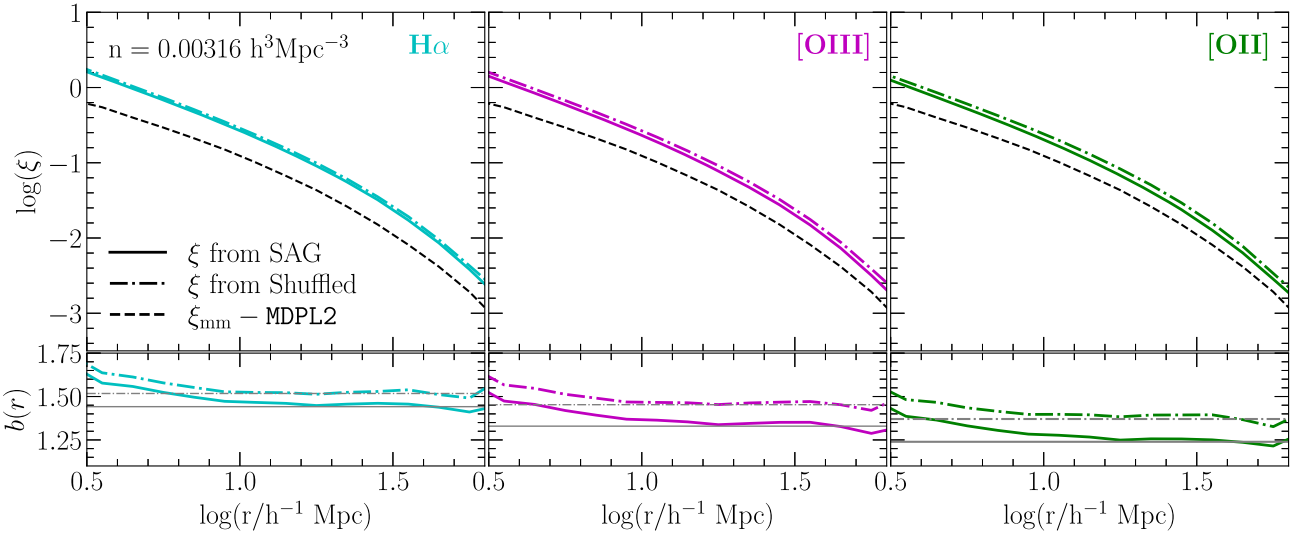

Figure 6. (*Top*) The 2PCFs of the SAG samples (solid) and the shuffled samples (dashed–dotted) for the Hα (*left-hand panel*), [O III] (*middle panel*), and [O II] (*right-hand panel*) selections. The black dashed lines correspond to the 2PCF of the dark matter distribution of the MDPL2 simulation. (*Bottom*) The bias parameter, as a function of scale separation, for the SAG and shuffled samples. The horizontal solid (dashed) grey line corresponds to the bias parameters of the SAG (shuffled) samples averaged between $1.4 < \log(r/h^{-1}\,\mathrm{Mpc}) < 1.8$.

technique. Therefore, our density-based definition and a cosmic web classification for environments might not be interchangeable.

We now consider the contribution of different haloes to the sample bias and their environment. Following Kim et al. (2009), we compute the effective clustering bias as a function of halo mass for each galaxy sample, and show the results in the top panel of Fig. 7. This parameter quantifies the contribution of galaxies in haloes of a given mass to the large-scale galaxy clustering amplitude of the sample. The effective bias is simply computed as $b(M) \times \Phi(M) \times <N(M)>$, where $b(M)$ is the bias function, $\Phi(M)$ is the halo mass function, and $<N(M)>$ is the halo occupation function of the galaxy sample. For each selection, the effective bias reaches a peak at different halo masses, close to the location of the 'knee' of the occupation function (i.e. when the highest fraction of haloes of a given mass contain a central). The middle panel shows the average $n_{\rm local}$ for the SAG and shuffled samples; the bottom panel shows the ratios between the $n_{\rm local}$ of the SAG and shuffled samples for each galaxy selection. Note that a ratio higher than unity indicates that galaxies in the SAG samples are in higher density regions than their shuffled counterparts.

The SAG galaxy samples exhibit different average $n_{\rm local}$, which suggests that these different galaxy populations live in different environments. In contrast, the averaged $n_{\rm local}$ for the shuffled samples are largely the same for all selections. Therefore, it appears that the shuffling procedure removes correlations between the selection and environment. For example, for the [O II]-selected galaxies from SAG, the shuffling removes the environmental preference for [O II] emitters to reside in low-density regions. As the shuffling procedure moves galaxies between haloes of the same mass, the resulting HODs of the shuffled samples are almost identical to those of their SAG counterparts, which in turn are notably different between selections (see Fig. 3). Thus, their $b(M)$ and 2PCFs are also different even when their averaged $n_{\rm local}$ are similar.

For the stellar mass-selected sample, we see that the ratio of $n_{\rm local}$ in the SAG model to that in the shuffled counterpart is close to unity, which indicates that the shuffling procedure does not modify the environment for this particular selection. For the SFR- and ELG-selected samples, on the other hand, the ratios in the bottom panel of Fig. 7 show clear departures from unity. Moreover, the $n_{\rm SAM}/n_{\rm Shuffle}$ ratios (i.e. environmental modifications due to the shuffling procedure) depend on halo mass. The resulting impact on the galaxy clustering, if it is present, should come from haloes whose masses are close to the peak of the effective clustering bias. For

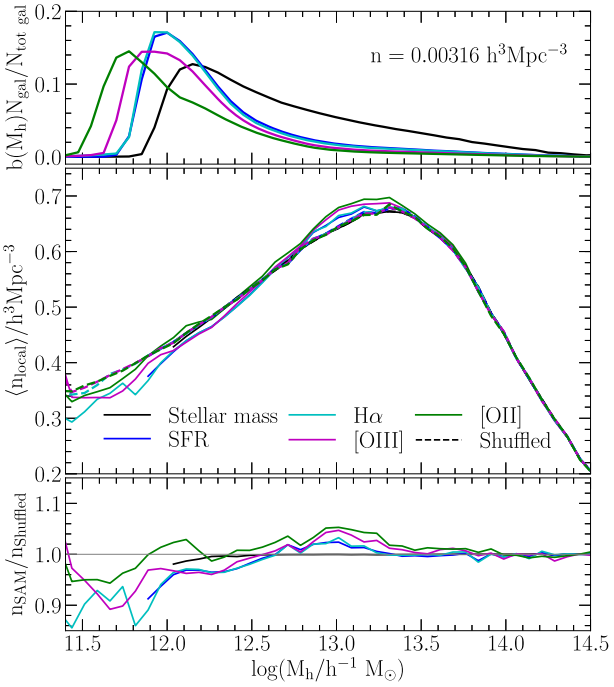

Figure 7. (*Top*) The contribution to the effective clustering bias from haloes as a function of halo mass. (*Middle*) The average local number densities for the SAG samples (solid) and their shuffled samples (dashed) as a function of halo mass, defined as described in the text. (*Bottom*) The ratio between the SAG and shuffled density measurements. Different colours indicate different galaxy selections as labelled in the middle panel.

Downloaded from https://academic.oup.com/mnras/article/506/3/3155/6313299 by Durham University Library user on 19 August 2021

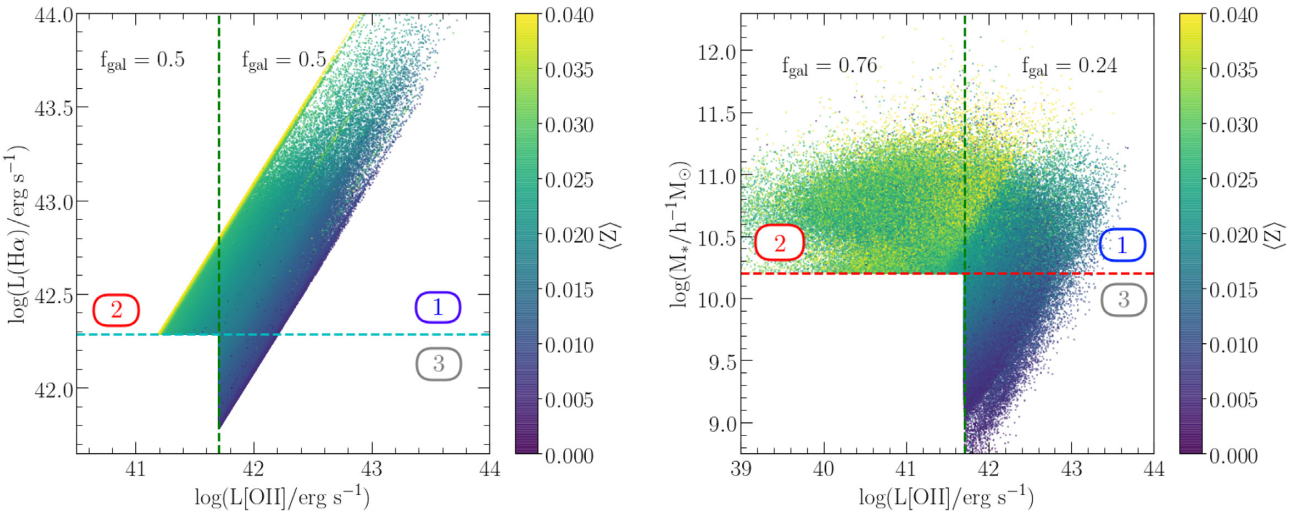

Figure 8. (*Left*) The H α emission as a function of [O II] emission colour coded by the cold mass-weighted metallicity of the discs and bulges. The cyan (green) dashed line indicates the cut in H α ([O II]) for a sample of number density $n = 0.003\,16\ h^3\ \mathrm{Mpc}^{-3}$. The fraction of galaxies with [O II] emission below and above the [O II] cut are included in both sectors. (*Right*) Same as the left but for stellar mass as a function of [O II] emission. The coloured numbers are labels for the quadrants in which galaxies lie, and are referred to in Fig. 9.

example, for the [O II] selection, the overlap between the peak of the effective bias and the mass dependence of the density ratio occurs in the halo mass range of $11.6 < \log(M_\mathrm{h}/h^{-1}\,\mathrm{M}_\odot) < 12$. In contrast, for the SFR and H α selections, we find no dependence on halo mass around the peak of the effective bias in $\log(M_\mathrm{h}/h^{-1}\,\mathrm{M}_\odot) \sim 12$. For the [O III]-selected sample, the ratio also depends on halo mass but in a narrower range of $11.7 < \log(M_\mathrm{h}/h^{-1}\,\mathrm{M}_\odot) < 11.9$, which is close to the peak of the effective bias.

Another approach to explore the origin of the scale-dependent assembly bias signature is to analyse the distribution of the gas-phase metallicity in the SAG samples. As the GET_EMLINES code uses this property as an input to predict the [O III] and [O II] emission-line luminosities, we expect that it is correlated to some extent with the spatial distribution of the [O III] and [O II] selections. Fig. 8 shows the distributions of SAG galaxies in the L(H α)–L[O II] and stellar mass–L[O II] planes. The points are colour coded by metallicity-averaged weighting by the mass of cold gas in both discs and bulges. We divide the galaxies into three subsamples separated by cuts in stellar mass, L(H α), and L[O II] (dashed lines) that correspond to a number density $n = 0.003\,16\ h^3\ \mathrm{Mpc}^{-3}$. Galaxies to the right (above) of the vertical (horizontal) dashed lines are contained in the [O II]-selected sample (stellar mass and H α selections). In this way, the overlap between the galaxy samples can be easily seen; for the L(H α)–L[O II] plane, we see that about half of [O II]-selected galaxies are contained in the H α selection, while for the stellar mass–L[O II] comparison, the [O II] emitters account for 24 per cent of galaxies in the stellar mass selection. Moreover, galaxies in the [O II] selection tend to be more metal poor than their H α or stellar mass counterparts, which is consistent with the metallicity distributions in Fig. 4. Indeed, it is clear that a large fraction of the [O II]-selected galaxies, in the bottom-right sectors, are the most metal poor.

We compute the autocorrelation functions of galaxies in each quadrant of the stellar mass–L[O II] and L(H α)–L[O II] planes to look for interesting features in their spatial clustering. We apply the shuffling procedure to these subsamples and measure the 2PCF of the resulting shuffled samples. We also compute the ratios of the 2PCF measured from the SAG subsamples to those of their shuffled counterparts to obtain the assembly bias signatures. The top panels in Fig. 9 show the assembly bias signatures for each subsample, colour coded by its quadrant in the two planes in Fig. 8. There is a remarkable difference between the assembly bias signatures of each quadrant; the red-coded galaxies, which are not included in the [O II] selections, show almost constant assembly bias, whereas the blue-coded ones show a prominent scale dependence. Moreover, the assembly bias for the grey-coded quadrants (which mostly contains metal-poor galaxies) exhibits a steeper dependence on separation. These two results suggest that galaxies with low gas-phase metallicity are driving the scale-dependent assembly bias. To connect this information with the environment of host haloes, we show, in the bottom panels of Fig. 9, the distribution of local number densities for galaxies in each quadrant. There is a subtle preference for grey-coded galaxies to live in less dense regions than galaxies in the other quadrants. These results suggest that the gas-phase metallicity is the property of the [O III] and [O II] selections that produces the scale-dependent assembly bias. Specifically, galaxies with low metallicity, which live in low-density regions, appear to account for most of the scale dependence. Nevertheless, further studies are needed to fully understand the correlation between metallicity and the spatial distribution of galaxies.

6 IMPACT ON COSMOLOGY

In the previous sections, we analysed the scale-dependent assembly bias in the [O III]- and [O II]-selected samples and its relation to the gas-phase metallicity and the environment of the galaxies. The scale dependence was found to be driven by low-metallicity galaxies in underdense environments. The scale dependence may have important implications for cosmological analyses.

In this section, we focus on the BAO feature and the β parameter that describes RSD. These quantities are measured from each of the SAG samples to check if the ELG selection introduces any systematic effects into the inference of cosmological parameters. This analysis is particularly important for the [O III] and [O II] selections, as they

Downloaded from https://academic.oup.com/mnras/article/506/3/3155/6313299 by Durham University Library user on 19 August 2021

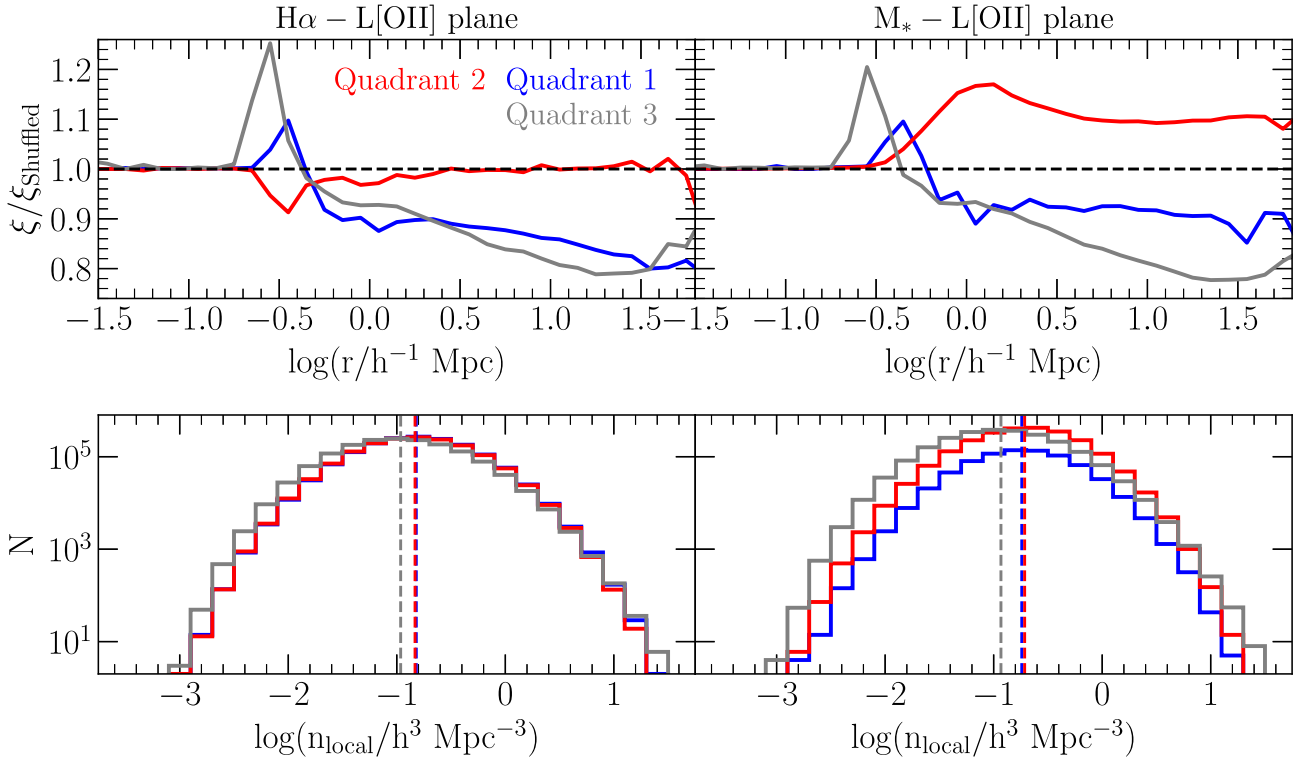

Figure 9. (*Top*) The assembly bias signatures for galaxies in the first (blue), second (red), and third (grey) quadrants of the Hα–*L*[O II] (left) and stellar mass–*L*[O II] (right) planes in Fig. 8. (*Bottom*) The local density distribution for galaxies in the corresponding quadrants. The dashed vertical lines correspond to the mean of each distribution.

contain particular features such as a non-constant bias and a galaxy assembly bias signal driven by the environment of these galaxies.

Fig. 10 shows the 2PCFs of the SAG (top panels) and shuffled samples (bottom panels) for different selection criteria and number densities. Note that, in order to focus on the BAO peak, we display $r^2 \times \xi(r)$. For comparison, we show the $z \sim 1$ linear theory prediction for the dark matter 2PCF of the MDPL2. The vertical dotted lines mark the position of the BAO peak for the dark matter.

The amplitudes of the BAO peaks are affected by the large-scale bias of each sample, which depends on the selection criteria, number density, and the degree of assembly bias present in the samples (recall the wide range of values found for the large-scale bias for the SAG and shuffled samples in Fig. 6). To estimate the position of the BAO peak, for each selection, we grid the simulation box into cubes of 125 h^{-1} Mpc a side, resulting in a total of 512 subvolumes. We compute the 2PCFs of the corresponding jackknife subsamples and fit a third-order polynomial to each to obtain the BAO peak scale, $r_{\rm peak}$. The final location of the BAO peak and its error are obtained from the mean and scatter of the individual jackknife measurements, respectively.[5] These are shown as the filled circles and error bars for each selection in Fig. 10.

Overall, the BAO peak scales are in close agreement with the linear theory prediction. Moreover, there are no significant differences between results from the SAG and shuffled samples. However, it is interesting to note that the shape of the BAO peak for the SAG [O II]-selected sample with the intermediate number density (top-middle panel) is slightly different relative to the other cases. Still, the large errors for all selections and number densities make it hard to claim a departure from the $r_{\rm peak}$ for the dark matter. A larger simulation is required to confirm the significance of this feature and whether it is related to the scale-dependent assembly bias. Nevertheless, it is still unclear whether such a shift would result in a biased measurement of cosmological parameters, as the environmental effects that influence the clustering measurements for the [O II] sample could also be interpreted as a non-local bias, which can be introduced as a nuisance parameter in BAO peak analyses (e.g. Sánchez, Baugh & Angulo 2008).

We also compute the β parameter for the SAG samples. This parameter is a function of the logarithmic growth rate, which depends on the matter density parameter, and the bias parameter of the galaxy sample. Following Padilla et al. (2019), we compute β for shuffled samples where the relative velocities of the galaxies within haloes – in addition to their positions – are maintained when shuffling satellites between haloes. The β parameter can be obtained from the ratio between the monopoles of the correlation functions in real and redshift space (Kaiser 1986b):

$$\xi_0(s) = \left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right)\xi(r). \quad (2)$$

The main panel of Fig. 11 shows the β parameter, calculated from equation (2), as a function of scale for the SAG (solid) and shuffled samples (dashed) with a number density of $n = 0.003\,16\ h^3\,{\rm Mpc}^{-3}$. For the SAG samples, we see that the β parameter is roughly constant for the stellar mass selection, and it is scale dependent for the other selections. However, for the shuffled samples, the scale dependence of β remains. The bottom panel of Fig. 11 shows the ratios between the β parameter of the SAG samples to that of their

[5]We also fit second-degree polynomials and Gaussian functions to the jackknife samples and find almost identical results as using the third-degree polynomials.

Downloaded from https://academic.oup.com/mnras/article/506/3/3155/6313299 by Durham University Library user on 19 August 2021

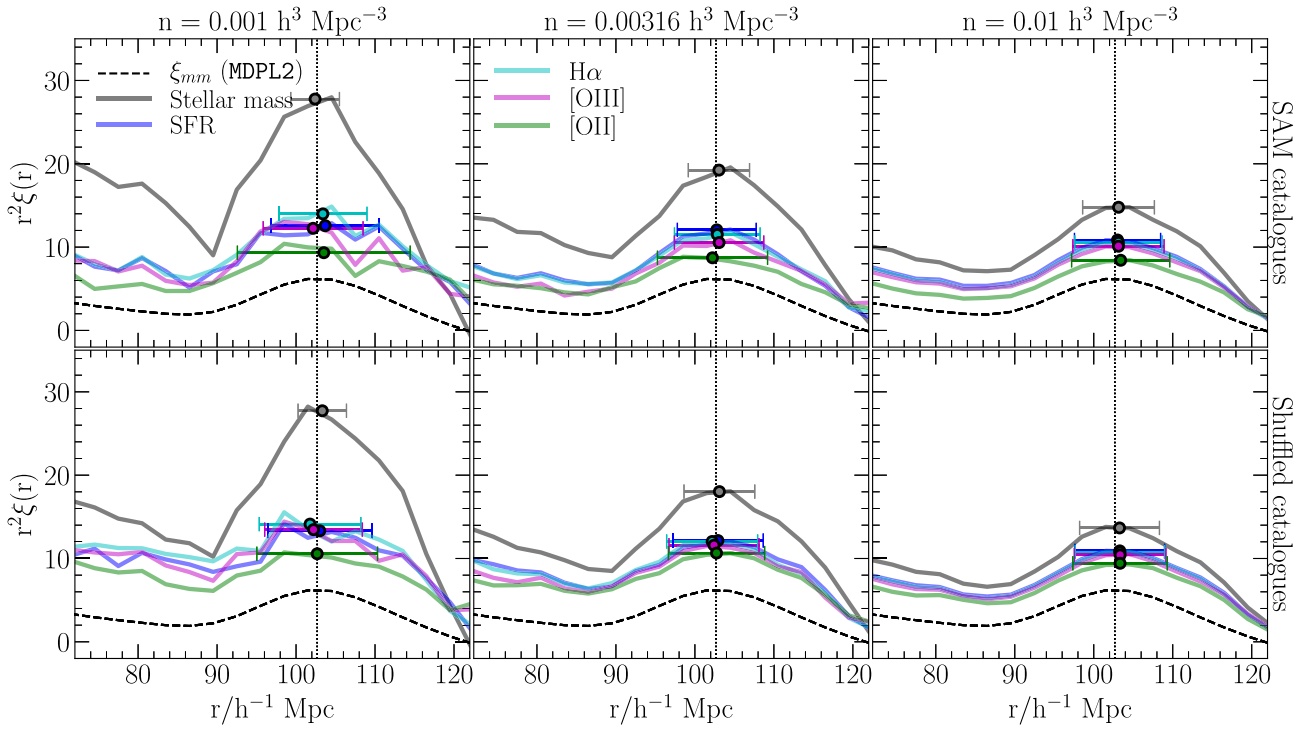

Figure 10. The BAO feature for the SAG samples (*top*) and their shuffled counterparts (*bottom*) with number densities of $n = 0.001\ h^3\,\mathrm{Mpc}^{-3}$ (*left*), $n = 0.003\,16\ h^3\,\mathrm{Mpc}^{-3}$ (*middle*), and $n = 0.01\ h^3\,\mathrm{Mpc}^{-3}$ (*right*). Colours correspond to the different selections adopted, as shown in the keys. The vertical lines are included to guide the eye and indicate the position of the BAO peak for the dark matter, according to the linear theory prediction (black dotted line). Filled points and error bars correspond to the scales of the BAO peak positions and their errors obtained from 512 jackknife subsamples (see the text for details).

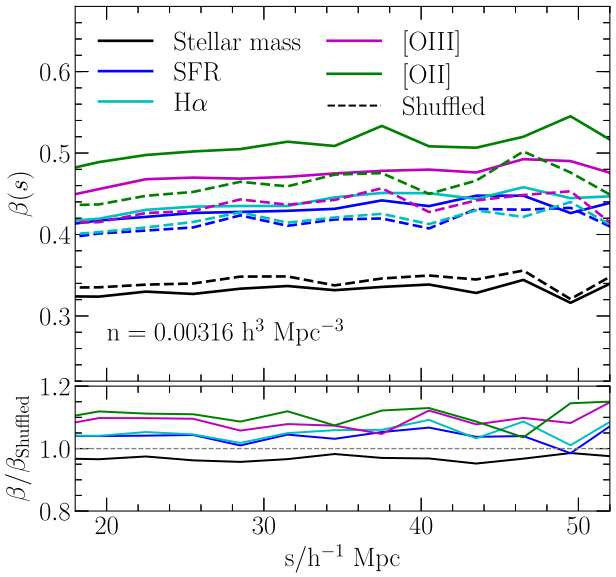

Figure 11. (*Top*) The β parameter as a function of separation for the SAG (solid) and shuffled samples (dotted) with a number density $n = 0.003\,16\ h^3\,\mathrm{Mpc}^{-3}$. Different colours correspond to different selection criteria as indicated by the keys. (*Bottom*) The ratio between the β parameters of the SAG and shuffled samples. The grey dashed line indicates unit ratio.

shuffled counterparts. For the stellar mass selection, the ratio is almost constant for all scales. For SFR- and ELG-selected samples, the ratio is more affected by noise, but it appears to be consistent with a constant value, with the SFR- and ELG-selected samples returning a higher value of beta than their shuffled counterparts. In principle, the steepness of β could be so slight that it falls below our noise level, even though we do find a scale-dependent bias for the ELG SAG samples, and a roughly constant one for the shuffled catalogues, especially for the [O III] and [O II] selections (see Fig. 6).

7 CONCLUSIONS

The next generation of galaxy surveys such as DESI and *Euclid* will map the sky by measuring redshifts to unprecedented numbers of ELGs. To fully exploit these upcoming data, we need to understand how these galaxies trace the underlying density of the Universe and to establish if there are any systematic effects that might impair our ability to extract unbiased cosmological information.

To investigate the potential of ELGs to constrain cosmological parameters, we study the clustering and halo occupation of the model galaxies from the SAG SAM (Cora et al. 2018) applied to the MDPL2 simulation outputs (Klypin et al. 2016). We use the instantaneous SFR and gas-phase metallicities of galaxies as the inputs to the GET_EMLINES code to obtain the nebular line emission of the galaxies. To mimic the selection criteria of future surveys, we use fixed number density samples where galaxies are ranked according to their [O III], [O II], Hα, and Hα+[N II] line luminosities. For comparison, we also include SFR- and stellar mass-selected samples.

We measure the 2PCFs, the HODs, and the galaxy assembly bias signatures for each galaxy sample. The galaxy assembly bias is measured via the ratio between the 2PCF of a SAG sample with that of a shuffled version of the sample, which, by construction, does not contain assembly bias. We also compute the absolute large-scale bias of ELGs to look for correlations between assembly bias, large-scale bias, the environment of the galaxies, and the gas-phase

Downloaded from https://academic.oup.com/mnras/article/506/3/3155/6313299 by Durham University Library user on 19 August 2021

metallicity of the ELG-selected samples. Finally, we measure the BAO feature and the β parameter for the SAG and shuffled samples to investigate the implications for cosmological studies using ELGs. Our results can be summarized as follows:

(i) ELG-selected samples have 2PCFs and HODs that are similar to those of SFR-selected galaxies. However, the [O III]- and [O II]-selected samples are less clustered than either the SFR or Hα samples. Moreover, their HODs indicate that most of the selected galaxies live in low-mass haloes. These differences explain why selecting by the luminosity of the [O III] or [O II] lines does not reproduce the behaviour of an SFR-selected sample.

(ii) The assembly bias signature (i.e. the ratio between the 2PCFs measured for the SAG and shuffled samples) for the [O III]- and [O II]-selected samples is scale dependent, with a steepness that becomes more pronounced for lower density samples (higher [O III] and [O II] thresholds). For the SFR and Hα selections, the assembly bias is scale dependent for samples with the lowest number density. In the case of galaxies selected by stellar mass, the assembly bias is roughly constant for all number densities.

(iii) The large-scale bias, defined as the ratio between the 2PCF of a galaxy sample and that of the dark matter, is scale dependent for the [O III]- and [O II]-selected samples in the SAG model. For the shuffled samples, in contrast, the large-scale bias is roughly constant for all selections. This suggests that the shuffling procedure removes an encoded dependence between the galaxy properties of [O III] and [O II] selections with the environment.

(iv) For a fixed halo mass, the local number density – which we use to quantify the environment of host haloes – is roughly the same for all shuffled samples. This indicates that the shuffling procedure eliminates the correlation between the selection and the environment of host haloes. In contrast, for the SAG samples the local number densities are notably different between the selection criteria, which indicates that in some cases the selected galaxies live in special environments. Moreover, the change of the environment of the [O II]-selected sample has a strong dependence with halo mass (see Fig. 7).

(v) Galaxies with low gas-phase metallicities are the ones that produce the scale-dependent assembly bias signature. Indeed, for the SAG sample, a larger fraction of metal-poor galaxies results in a steeper scale-dependent assembly bias signature (see Fig. 9). Moreover, these galaxies tend to live in low-density regions, which is more common for the ELG-selected samples.

(vi) The BAO peak scale for all selections and number densities is in close agreement with that for the dark matter. Interestingly, the BAO peak for the SAG [O II] selection with a number density $n = 0.003\,16\ h^3\,\mathrm{Mpc}^{-3}$ has a slightly different shape than all other cases. An analysis on a larger simulation should be done to confirm whether this feature in the BAO peak is important.

(vii) The β parameter for the SFR- and ELG-selected samples is non-constant as a function of scale for the SAG and shuffled samples. This is clearer for the [O II] and [O III] selections and can be explained as a combination of the scale-dependent large-scale bias and a possible non-constant logarithmic growth rate. For the stellar mass case, in contrast, β is roughly constant.

Our results show that care must be given when using future galaxy samples from *Euclid* and DESI, which will be selected by their emission-line luminosities. We find that this type of selection can produce samples that lie in special environments, and because of this their clustering can show a different slope than that of the underlying matter density field. This type of environment selection needs to be modelled and marginalized over in cosmological parameter constraints from such samples in order to avoid systematic effects in their analysis.

ACKNOWLEDGEMENTS

We thank the anonymous referee for the insightful comments that helped us to improve the quality of this paper. This work was made possible by the efforts of Gerard Lemson and colleagues at the German Astronomical Virtual Observatory in setting up the Millennium Simulation data base in Garching. The CosmoSim data base used in this paper is a service by the Leibniz-Institute for Astrophysics Potsdam (AIP). The MultiDark data base was developed in cooperation with the Spanish MultiDark Consolider Project CSD2009-00064. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) and the Partnership for Advanced Supercomputing in Europe (PRACE; http://www.prace-ri.eu) for funding the MultiDark simulation project by providing computing time on the GCS Supercomputer SuperMUC at Leibniz Supercomputing Centre (LRZ; http://www.lrz.de). EJ and NP acknowledge support from 'Centro de Astronomía y Tecnologías Afines' BASAL AFB-170002 and ANID FONDECYT Regular 1191813. IZ acknowledges support by NSF grant AST-1612085. This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement number 734374. The calculations for this paper were performed on the Geryon computer at the Center for Astro-Engineering UC, part of the BASAL PFB-06, which received additional funding from QUIMAL 130008 and Fondequip AIC-57 for upgrades.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Alam S. et al., 2021, Phys. Rev. D, 103, 083533
Artale M. C., Zehavi I., Contreras S., Norberg P., 2018, MNRAS, 480, 3978
Avila S. et al., 2020, MNRAS, 499, 5486
Baldwin J. A., Phillips M. M., Terlevich R., 1981, PASP, 93, 5
Baugh C. M., 2006, Rep. Prog. Phys., 69, 3101
Baugh C. M. et al., 2019, MNRAS, 483, 4922
Bautista J. E. et al., 2018, ApJ, 863, 110
Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013a, ApJ, 762, 109
Behroozi P. S., Wechsler R. H., Wu H.-Y., Busha M. T., Klypin A. A., Primack J. R., 2013b, ApJ, 763, 18
Benson A. J., 2012, New Astron., 17, 175
Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000, MNRAS, 311, 793
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Byler N., Dalcanton J. J., Conroy C., Johnson B. D., 2017, ApJ, 840, 44
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Cole S. et al., 2005, MNRAS, 362, 505
Contreras S., Baugh C. M., Norberg P., Padilla N., 2013, MNRAS, 432, 2717
Contreras S., Zehavi I., Padilla N., Baugh C. M., Jiménez E., Lacerna I., 2019, MNRAS, 484, 1133
Contreras S., Angulo R. E., Zennaro M., 2021, MNRAS, 504, 5205
Cora S. A., 2006, MNRAS, 368, 1540
Cora S. A. et al., 2018, MNRAS, 479, 2
Croton D. J. et al., 2006, MNRAS, 365, 11
Croton D. J., Gao L., White S. D. M., 2007, MNRAS, 374, 1303
Croton D. J. et al., 2016, ApJS, 222, 22
Dawson K. S. et al., 2013, AJ, 145, 10

Downloaded from https://academic.oup.com/mnras/article/506/3/3155/6313299 by Durham University Library user on 19 August 2021

De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
De Lucia G., Kauffmann G., White S. D. M., 2004, MNRAS, 349, 1101
DESI Collaboration, 2016, preprint (arXiv:1611.00036)
Dopita M. A., Sutherland R. S., 1995, ApJ, 455, 468
Eisenstein D. J. et al., 2005, ApJ, 633, 560
Eisenstein D. J. et al., 2011, AJ, 142, 72
Favole G. et al., 2016, MNRAS, 461, 3421
Favole G. et al., 2020, MNRAS, 497, 5432
Gao L., White S. D. M., 2007, MNRAS, 377, L5
Gao L., Springel V., White S. D. M., 2005, MNRAS, 363, L66
Gargiulo I. D. et al., 2015, MNRAS, 446, 3820
Gonzalez-Perez V. et al., 2018, MNRAS, 474, 4024
Gonzalez-Perez V. et al., 2020, MNRAS, 498, 1852
Groves B. A., Allen M. G., 2010, New Astron., 15, 614
Groves B. A., Dopita M. A., Sutherland R. S., 2004, ApJS, 153, 9
Guo Q. et al., 2011, MNRAS, 413, 101
Guo Q., White S., Angulo R. E., Henriques B., Lemson G., Boylan-Kolchin M., Thomas P., Short C., 2013, MNRAS, 428, 1351 (G13)
Guo H. et al., 2019, ApJ, 871, 147
Gutkin J., Charlot S., Bruzual G., 2016, MNRAS, 462, 1757
Hearin A. P., Watson D. F., van den Bosch F. C., 2015, MNRAS, 452, 1958
Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R. E., Guo Q., Lemson G., Springel V., 2013, MNRAS, 431, 3373
Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, MNRAS, 451, 2663
Jiménez E., Contreras S., Padilla N., Zehavi I., Baugh C. M., Gonzalez-Perez V., 2019, MNRAS, 490, 3532
Kaiser N., 1986a, MNRAS, 219, 785
Kaiser N., 1986b, MNRAS, 222, 323
Kim H.-S., Baugh C. M., Cole S., Frenk C. S., Benson A. J., 2009, MNRAS, 400, 1527
Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340
Knebe A. et al., 2018, MNRAS, 474, 5206
Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, ApJ, 609, 35
Lacerna I., Padilla N., Stasyszyn F., 2014, MNRAS, 443, 3107
Lagos C. D. P., Cora S. A., Padilla N. D., 2008, MNRAS, 388, 587
Lagos C. d. P., Tobar R. J., Robotham A. S. G., Obreschkow D., Mitchell P. D., Power C., Elahi P. J., 2018, MNRAS, 481, 3573
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Levesque E. M., Kewley L. J., Larson K. L., 2010, AJ, 139, 712
Lin Y.-T., Mandelbaum R., Huang Y.-H., Huang H.-J., Dalal N., Diemer B., Jian H.-Y., Kravtsov A., 2016, ApJ, 819, 119
Merson A., Wang Y., Benson A., Faisst A., Masters D., Kiessling A., Rhodes J., 2018, MNRAS, 474, 177
Montero-Dorta A. D. et al., 2017, ApJ, 848, L2
Montero-Dorta A. D. et al., 2020, MNRAS, 496, 1182
Muñoz Arancibia A. M., Navarrete F. P., Padilla N. D., Cora S. A., Gawiser E., Kurczynski P., Ruiz A. N., 2015, MNRAS, 446, 2291
Nagao T., Maiolino R., Marconi A., 2006, A&A, 459, 85
Nelson D. et al., 2018, MNRAS, 475, 624
Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, MNRAS, 396, 19
Orsi Á., Padilla N., Groves B., Cora S., Tecce T., Gargiulo I., Ruiz A., 2014, MNRAS, 443, 799 (O14)
Padilla N. D., Salazar-Albornoz S., Contreras S., Cora S. A., Ruiz A. N., 2014, MNRAS, 443, 2801
Padilla N., Contreras S., Zehavi I., Baugh C. M., Norberg P., 2019, MNRAS, 486, 582
Planck Collaboration XVI, 2014, A&A, 571, A16
Ruiz A. N. et al., 2015, ApJ, 801, 139
Sánchez A. G., Baugh C. M., Angulo R. E., 2008, MNRAS, 390, 1470
Schaye J. et al., 2015, MNRAS, 446, 521
Scoccimarro R., Feldman H. A., Fry J. N., Frieman J. A., 2001, ApJ, 546, 652
Shim H., Chary R.-R., 2013, ApJ, 765, 26
Sinha M., Garrison L. H., 2020, MNRAS, 491, 3022
Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008, MNRAS, 391, 481
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Springel V. et al., 2005, Nature, 435, 629
Stevens A. R. H., Croton D. J., Mutch S. J., 2016, MNRAS, 461, 859
Tecce T. E., Cora S. A., Tissera P. B., Abadi M. G., Lagos C. D. P., 2010, MNRAS, 408, 2008
Vogelsberger M. et al., 2014, MNRAS, 444, 1518
Walsh K., Tinker J., 2019, MNRAS, 488, 470
Watson D. F. et al., 2015, MNRAS, 446, 651
Wechsler R. H., Tinker J. L., 2018, ARA&A, 56, 435
Wechsler R. H., Zentner A. R., Bullock J. S., Kravtsov A. V., Allgood B., 2006, ApJ, 652, 71
Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, Phys. Rep., 530, 87
Xu X., Zheng Z., 2020, MNRAS, 492, 2739
Xu X., Zehavi I., Contreras S., 2021, MNRAS, 502, 3242
Zehavi I. et al., 2011, ApJ, 736, 59
Zehavi I., Contreras S., Padilla N., Smith N. J., Baugh C. M., Norberg P., 2018, ApJ, 853, 84
Zehavi I., Kerby S. E., Contreras S., Jiménez E., Padilla N., Baugh C. M., 2019, ApJ, 887, 17
Zheng Z. et al., 2005, ApJ, 633, 791
Zu Y., Mandelbaum R., 2016, MNRAS, 457, 4360

APPENDIX A: ASSEMBLY BIAS OF ELGS IN L-GALAXIES

To assess if the scale-dependent assembly bias can be found in other SAMs, we select galaxy samples at $z = 0$ from the Guo et al. (2013) model (hereafter G13) that is a version of the L-GALAXIES code from the Munich group (De Lucia, Kauffmann & White 2004; Croton et al. 2006; De Lucia & Blaizot 2007; Guo et al. 2011; Henriques et al. 2013). G13 is a semi-analytic model, and as such it models a set of physical processes that shape the formation and evolution of galaxies, applied to halo merger trees drawn from the Millennium-WMAP7 simulation. This simulation was carried out in a box of 500 h^{-1} Mpc a side, and is the same as the original Millennium simulation (Springel et al. 2005) but with updated cosmological parameters that match the results from the WMAP7 observations.

We use the GET_EMLINES code to obtain the nebular emission for G13 galaxies. The instantaneous SFR is not a direct output of G13; hence, we use the *average* SFR instead to infer line emission luminosities. This is motivated by the results of Favole et al. (2020); they demonstrate that using average SFRs as inputs for GET_EMLINES produces good predictions to study average populations of [O II] ELGs. We then define new stellar mass-, SFR-, Hα-, [O III]-, and [O II]-selected samples following the procedure given in Section 2.4.

Fig. A1 shows the assembly bias signatures for the G13 samples with three different number densities (the same ones used to define the SAG samples). Note that the 2PCFs of the Hα and SFR selections are the same, which is a consequence of Hα luminosity having a simple dependence on the SFR, with little variation with the cold gas metallicity. Even though the clustering measurements are noisier at very large scales, we still find that the [O III] and [O II] selections have a clear scale-dependent assembly bias. In contrast, for the SFR and Hα selections, the scale dependence is there for the lowest number density sample alone, while for the stellar mass-selected samples the signature is roughly flat in all cases. The shaded regions correspond to the uncertainties estimated from jackknife resampling using 125 cubic volumes of 100 h^{-1} Mpc side. Note that due to the smaller volume of the Millennium simulation, the 2PCFs are overall noisy and not well defined for $\log(r/h^{-1}\,\mathrm{Mpc}) > 1.5$.

Downloaded from https://academic.oup.com/mnras/article/506/3/3155/6313299 by Durham University Library user on 19 August 2021

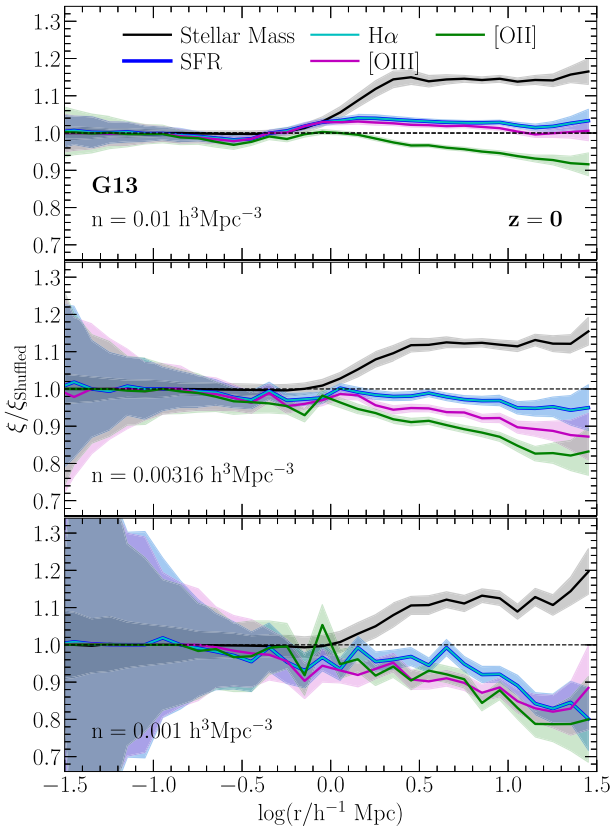

Figure A1. Same as the left-hand panels of Fig. 5 but for galaxy samples extracted from the Guo et al. (2013) SAM. Note that due to the smaller size of the simulation box we show results up to $\log(r/h^{-1}\,\mathrm{Mpc}) = 1.5$.

APPENDIX B: COMPLETENESS OF ELG-SELECTED SAMPLES

Selecting by emission-line luminosities produces samples that trace not only the amplitude of SFR but also other additional properties, such as the cold gas metallicity. Thus, in principle, a low-SFR galaxy may be included in an ELG-selected sample. Because of this, we analyse the effect of the moderate stellar mass cut imposed on the SAG data, which is present in the subsamples analysed in this work; this moderate cut is $M_* > 10^{8.7}\,h^{-1}\,\mathrm{M}_\odot$, which is slightly lower than the resolution of the MDPL2 and Millennium simulations ($\sim 10^9\,h^{-1}\,\mathrm{M}_\odot$).

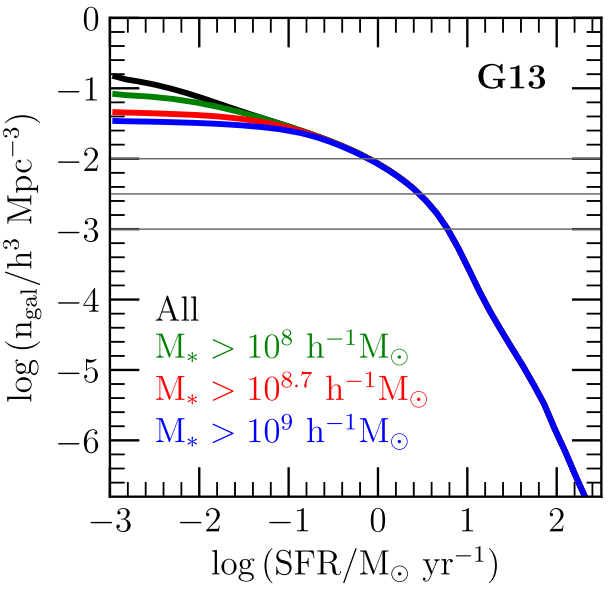

Figure B1. The cumulative SFR function of subsamples of the Guo et al. (2013) model defined by different stellar mass cuts indicated by the colours and labels. Horizontal lines indicate the different number densities used to define the galaxy samples.

As SAG and G13 show similar trends for assembly bias (see Fig. A1), we expect that the effect of the completeness stellar mass cut on these trends should be also similar for both models. Fig. B1 displays the cumulative SFR function for subsamples of G13, defined by different stellar mass cuts. As expected from the stellar mass–SFR relation, we see that the larger the cut, the smaller the number of low-SFR galaxies.

We define a subsample of G13 by selecting galaxies with stellar mass above the cut imposed for SAG. Then, following the procedure in Section 2.4, we define our galaxy samples with this new cut. We measure the assembly bias signatures for these samples, and we compare them with the assembly bias of G13 samples with no previous cuts in stellar mass. We find that the assembly bias signatures are almost identical for all selections and number densities. Noticeable differences only arise at very large scales [$\log(r/h^{-1}\,\mathrm{Mpc}) > 1.5$]. This indicates that the moderate stellar mass cut introduced to impose completeness has only a minor impact on the assembly bias signature, which includes the scale dependence for [O III] and [O II] selections.

This paper has been typeset from a TEX/LATEX file prepared by the author.

Downloaded from https://academic.oup.com/mnras/article/506/3/3155/6313299 by Durham University Library user on 19 August 2021